Comments on the ICCVAM Expert Panel Report on the IRE BRD

Provided by the Member of the COLIPA/SCAAT Eye Irritation Task Force:
Penny Jones (Unilever, Sharnbrook, UK)

- I am pleased that the Expert Panel has concluded that the IRE BRD proposed test method appears to be capable of identifying severe eye irritants. They state that this does need to corroborated with a larger number of test substances (i.e., it has not fully met ICCVAM validation criteria), but I am not sure it is entirely clear as to how they intend to get data that has been obtained using the proposed method, as most of the old data (with the exception of that from SafePharm which they have analysed) is not using this exact method.
- In 12.3 they state that there is no need to conduct any optimisation or validation studies until the existing available data has been analysed more thoroughly which I agree with. They concede (7.4) that the recommended protocol may be insensitive to minor changes and that results from other existing studies could therefore be used as validation data. I think they should definitely explore this, as this is the only way to get data from more materials for which the *in vivo* data is known.
- On the other hand they do leave the door open that more studies may need to be done (even possibly additional animal studies, which I don't think should be done). This may be the only way of getting interlab. reproducibility data for the particular protocol. I think it is going to be quite difficult for them to get any more studies done (even just *in vitro* studies) on an *ex vivo* assay like this, especially if they wanted several labs to take part, as I don't think people would be keen. I am not sure of the overall enthusiasm in the scientific community for further validation studies of an assay using rabbit eyes, when there may well be 3-D model assays available in the future.
- I am pleased that the Panel have made quite a lot of comment (4.6) on the lack of discussion in the BRD of the accuracy and reliability of the *in vivo* test and that the variability of the rabbit test should be taken into account of in any comparison of the *in vivo/in vitro* data. I take their point though that variability might not be so much of an issue at the severe end of the spectrum.
- The protocol recommended by the BRD has largely been accepted by the Panel with a few modifications. I think my original comments on the protocol still stand (I will attach the document I originally wrote in January 2005 on this as well to my e-mail sending these comments), accepting the fact that we do the assay slightly differently here at Unilever, but for other purposes than just C & L. The panel has also agreed with my original comment that the BRD should be more explicit that the recommended method is more or less the method of SafePharm. As the SafePharm method is used for screening for severe eye irritants it is logical that this should be the recommended method. It not clear whether the Panel have considered whether there may be any more data available via SafePharm using this method (this may have been done as SafePharm took part in the expert meeting)
- The Panel suggest inclusion of histopathology should be an option as an endpoint which I agree with (we do this routinely and not just for borderline results as they suggest) as this provides a permanent record of the cornea.

- The Panel have suggested that reference material like photographs for the mainly subjective endpoints should be available for training purposes which is a good idea (but might entail a lot of work on someone's part).
- It is suggested that the extent to which "leading edge" techniques could be used to measure corneal damage should be explored. This is a good idea in that anything which could convert subjective measurements into more objective numerical measurements would make scoring easier, but again I doubt whether anyone would want to spend a great deal of time and effort on doing this.
- The Panel discusses the source and availability of rabbit eyes for this assay (12.2.1). I agree with their comments that it needs to be ensured that appropriate sources are defined and that animals should not be sacrificed specifically to provide eyes for this test. They raise the question as to the availability of enough eyes if the assay were to be of widespread use, but they have not particularly discussed what might happen if supplies of eyes from animals used for other toxicological tests are not available.
- The Panel have discussed the decision criteria for identifying a severe irritant (2.1.12) (which are based on those used at SafePharm) and that there is not a rationale for them in the method. These must have been based on some sort of rationale, which ought to be available from SafePharm. They suggest that the criteria are tested out again the other studies available, but this would be difficult as the scoring for opacity and fluorescein staining would not have been done using the same system. This would need to be addressed in any further analysis of existing data (are the scoring systems sufficiently similar for comparison).
- There is discussion of the proposed list of reference materials, which were given in the BRDs. I agree with the comments of the Panel that there are too many materials (particularly the surfactant-based materials which were used in the CTFA studies). However, it should be possible to make a sensible list from the sets of single chemicals previously used for validation studies, for which the *in vivo* data should exist. A list for looking at the identification of severe irritants might be different though to a list for looking at the wide spectrum of irritation (list should be suitable for both?).
- At the end of the IRE BRD there is a minority report from one of the Panel (Dr Martin Stephens) and I would tend to agree with his comments.

Penny Jones, 21/04/2005

Comments on ICCVAM IRE Background Review Document

Overall, I welcome this review of the IRE and appreciate the amount of work put into the BRD so far by ICCVAM/NICEATM. The IRE BRD has two main components, the presentation of an "optimised" method for the test and the analysis of historic data, for which I would like to make the following comments:

"Optimised" method

Although the authors say that the method proposed is derived from all the protocols looked at it appears that it is essentially the protocol used by SafePharm, UK, (and in Guerriero et al, 2004) to identify severe eye irritants as part of a testing strategy for clients. This should be made clearer in the BRD. This is different in some aspects from the method used at Unilever, where the original method was developed (mainly in the method of dosing of liquids, use of McDonald-Shadduck scoring system, the actual prediction model used and use corneal histology). At Unilever, the IRE is currently used in a comparative mode as part of the product risk assessment process and not for classification of ingredients, and therefore I don't see the use of a specific protocol for identifying severe irritants as a problem, provided this does not undermine other uses of the assay.

The suggested prediction model is however slightly different from that which I understood SafePharm to currently use, in that it is suggested that the cut-off for Maximum Corneal Opacity should be 3, rather than 4, which I believed was the cut-off currently used. The BRD states that this is based on Guerriero et al, 2002. I had understood that the model currently used by SafePharm was statistically derived in collaboration with a client. The basis for the suggested prediction model should be clearly given in the BRD if possible to provide transparency and support for the validation process.

Routine corneal histology has not been recommended as an end-point, on the grounds that not all labs are equipped perform it; this is not a valid scientific reason for not doing it. Although perhaps histology is not essential for the identification of severe irritants it is a very useful permanent record of the treated/control cornea which should be recommended; it is comparatively easy for histology to be contracted out by labs which do not perform it themselves.

The method gives recommendations for the numbers of eyes per group to be used (3 per group for test and controls) and that positive and benchmark controls should be used as well as negative and solvent controls in each experiment. Although ideally one would accept this for a regulatory assay, the number of eyes which can be treated in any one experiment may limited in practical terms, e.g., the number of eyes that can be perfused at any one time, timing of treatment and measurements. This should be considered in the BRD.

Data analysis

The BRD has statistically reanalysed the data from four previous studies using the prediction model suggested in the "optimised" method comparing it to *in vivo* data reclassified according to EU, EPA and GHS systems. On the whole, I do not think that the minor differences in the methods compared with "optimised" protocol affect the overall results using the assay. However, there are some problems with using this prediction model with the data from three of the studies as not all the endpoints were used in all the studies and the scoring was not done using the McDonald-Shadduck scoring system (exception is Guerriero et al, 2004 which used the scoring system on which this prediction model is generally based). Because prediction models were not used in these studies it was therefore difficult in this respect, but it might be possible to go back to the participants in these studies to ascertain the scoring systems and models in use at the time. However, these problems are acknowledged in the BRD which states that it is impossible to pool the data from the studies because of this and that the most complete data available is from Guerriero et al, 2004, on which the prime conclusions of the BRD in terms of accuracy etc are based. The main interpretation from this statistical analysis are that severe irritants/corrosives would not be under predicted, but that there is quite a high rate of false positives. On the whole I would probably concur with this with our experience that materials are more likely to be over predicted than under predicted.
In the past the failure of the validation studies (reevaluated in this BRD) has in part been blamed on the *in vivo* data and its variability. Although classification systems have been used in this BRD rather than numerical MMAS scores the evaluation has still been against the same *in vivo* data and little attempt has been made to take this into account or discuss the impact on the evaluation. This should be done in some way.

It is very disappointing that the only overall conclusion from the BRD is that further optimisation and validation studies are required without which the IRE can't be used in a testing strategy to identify severe irritants, when it is clear that the assay is being used successfully for such purpose. This also undermines its current acceptance on a case-by-case basis by the EU for labelling of severe irritants. There has been no attempt to use a weight-of-evidence approach to evaluate some of the other studies/publications which are available and cited in the BRD (e.g., Koeter and Prinsen, 1985; Lewis et al, 1994; Price and Andrews, 1985 and the P&G submitted data). Although the data is not perhaps in the form that could be statistically analysed the usefulness of the assay in the context of the studies and the use to which the data was put at the time should be taken account of. Further investigation of existing data should be carried out before concluding that the only way to progress is yet another validation study.

<u>Other comments</u>

There are various statements in the BRD about the lack of GLP compliance of studies (apart from EC/HO study), however, I know that some of the other studies were performed in GLP compliant labs. Perhaps the authors should make more attempt to contact the original authors/participants for statements as to their GLP status at the time, although for most *in vitro* studies GLP compliance was not necessarily a requirement at the time and data from all competent labs should be considered.

Table 12.5 gives a table of recommended substances for use in validation studies. This includes formulations used in the CTFA study. Although these substances have *in vivo* data and the formulations available, I would query whether it would be possible to reproduce these formulations sufficiently well for them to be of use? It would be better to have well defined single chemicals.

Penny Jones, 06/01/05